DOCUMENT RESUME

ED 464 145 TM 033 829

AUTHOR Martinez, Mario C.; Ivory, Gary
TITLE The "Quality Counts 2002" State Report Cards: Correlations with Measures of State Education Environments.
PUB DATE 2002-04-00
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Correlation; *Demography; Educational Quality; Elementary Secondary Education; *Evaluation Utilization; *Policy Formation; Regression (Statistics); *Report Cards

ABSTRACT

Correlation and multiple regression analyses revealed the extent to which the report card grades of Quality Counts 2002, prepared by Education Week, were intercorrelated with and predicted by demographic factors among many states. The researchers used four grades from Quality Counts 2002 and five demographic factors. Few intercorrelations were found among Quality Counts 2002 grades, which suggested that the report cards are comprehensive. The multiple regression equations show that demographic factors account for from 4% to 66% of the variance in Quality Counts 2002 grades. This suggests that the report cards are moderately functional, i.e., they have potential to guide policy development. It is concluded that the Quality Counts report cards are useful for provoking discussion but not for resolving educational policy disputes. (Contains 2 tables and 32 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made

ED 464 145

Running Head: QUALITY COUNTS: RELATIONS WITH STATE ENVIRONMENTS

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M. Martinez

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

The *Quality Counts 2002* State Report Cards: Correlations with Measures of State Education Environments

Mario C. Martinez and Gary Ivory

New Mexico State University

Session 56.42: Using Data in Educational Decisions. American Educational Research Association Annual Meeting, New Orleans, LA, April 5, 2002

TM033829

BEST COPY AVAILABLE

Abstract

Correlation and multiple regression analyses revealed the extent to which *Education Week's Quality Counts 2002* report card grades were intercorrelated and predicted by demographic factors among states. The researchers used four grades from *Quality Counts 2002* and five demographic factors. The researchers found few intercorrelations among *Quality Counts 2002* grades, which suggested that the report cards are comprehensive. The multiple regression equations showed that demographic factors accounted for from four to 66 percent of the variance in *Quality Counts 2002* grades. This suggested that the report cards are moderately functional, i.e., they have potential to guide policy development. The researchers conclude that the Quality Counts report cards are useful for provoking discussion but not for resolving educational policy disputes.

The *Quality Counts 2002* State Report Cards: Correlations with Measures of State Education Environments

Accountability has long been an issue in K-12 education in the United States. Arguments on the pro side range from a Commissioner of Education's concern in the early 1960s that "American education had not yet faced up to the question of how to determine the quality of academic performance in the schools" and thus lacked the "support to shore up educational weakness" (Vinovskis, 1998, p. 5) to President George W. Bush's characterization of strong accountability measures as the best alternative to "the soft bigotry of low expectations" (White House, 2001). On the other side, opposition has ranged from the 1970 argument against the National Assessment of Educational Progress (NAEP) that evaluation is far more useful in local systems than on a national level (Vinovskis, 1998, p. 8) to McNeil's (2000) claims of the harm done by Texas's accountability system.

One form of accountability system is the state report card. In defense of organization report cards, Gormley and Weimer (1999, p. 21) argued that they can reduce "information asymmetry" (where service providers understand what they are doing better than the client) and thus enable both clients and public managers to pressure providers for better services. In fact, the Arizona Education Association (2002, p. 13) credited Arizona's ranking of 50th on *Education Week*'s *Quality Counts* report card for helping "build the momentum" for a sales tax increase to improve education there. But Mc'Donnell (1994) explained how thoroughly providers may disagree with public managers over the meaning of any given information. She also described the need policy makers feel to move "quickly while the policy window is open" (p. 401) and to base decisions on quality-of-education indicators despite their technical limitations.

On the negative side, detractors have argued that quality-of-education indicators are at worst harmful, and at best necessary but insufficient sources of guidance. Those who claim they are harmful cite practitioners' narrow focus on indicators and the fact that assessment systems cannot validly both enlighten practitioners and sanction them (McDonnell, 1994; McNeil, 2000) and their unintended negative consequences (Haney, cited in Scheurich, Skrla, & Johnson, 2000). Those who argue that quality-of-education indicators are necessary but insufficient guides claim that without attention to the process by which indicators and policy are translated into classroom practice (Knapp, 1997), to capacities of educational systems to change (Fullan, 2001), or to the environments in which they operate (Ginsberg & Berry, 1997), their adherents overlook critical issues.

Gormley and Weimer (1999) offered six criteria for judging report cards: validity, comprehensiveness, comprehensibility, relevance, reasonableness, and functionality. Two criteria of interest in our study, comprehensiveness and functionality, refer respectively to the facts that a report card should "be comprehensive in terms of important dimensions of organizational performance and . . .include a range of indicators" (p. 36) and that "a report card should be crafted in such a way that it convinces targeted organizations to engage in appropriate . . .behavior" (p. 37). Relevant to comprehensiveness, Ewell (2001) stated that subcategories that comprise a larger category should be neither too closely correlated nor too negatively correlated; a moderate correlation is best. High correlations would suggest that subcategories (e.g., report card grades) were assessing the same trait, and thus were redundant; negative correlations would suggest that high scores on one trait might actually work against achieving high scores on another trait. Kennedy (1999) categorized indicators of student outcomes and described the limitations of each category.

The visibility of state K-12 report cards has been elevated over the last seven years by the annual publication entitled *Quality Counts*, produced by *Education Week* (2002). In an analogous effort, the National Center for Public Policy in Higher Education (NCPPHE) (2000) produced higher education report cards for each state. *Quality Counts* provokes the question of the report cards' functionality, i.e., whether states' grades in K-12 education can inform policymakers' decisions. Are grades related to policymakers' decisions, or to demographic conditions that are beyond their control? Cunningham and Wellman (2001) addressed the identical question with regard to NCPPHE's higher education state report cards. Ackoff (1994) discussed the surroundings of an open system, and defined the contextual conditions of that system as those things that cannot be directly influenced or controlled. Richardson, Reeves-Bracco, Callan, and Finney (1999) described these factors as comprising the policy environment. Ackoff also defined variables that can be influenced or perhaps controlled as part of the transactional environment.

Scholars debate the relative weight of demographics and policy in influencing the quality of education. Two demographic factors often thought to influence education are wealth and ethnicity. With regard to wealth, Wenglinsky (1997) argued that since the Coleman report, sociologists tend to emphasize the importance of the environment over anything that schools can determine. As for ethnicity, differences between minority and majority standardized test scores are well documented (e.g., Borman, Stringfield, and Rachuba, 2000; College Board, 1999; Lee, 2002). Oakes and her associates (1990) revealed that the additive effects and interactions among income, minority status and school factors were complex and practically inseparable. In the policy arena, the research of Odden, Monk, Nakib, and Picus (1995) revealed little variation in how districts spend money. They claimed, therefore, that it was unsurprising that student

achievement did not correlate with spending patterns. They also argued that hiring larger numbers of teachers either to reduce class size or to provide more out-of-classroom services has not boosted student achievement very much. Wenglinsky (1997), however, claimed from his LISREL analysis of NAEP data and Common Core of Data for U.S. schools that spending money to reduce class size will result in higher student achievement. Finn (2002) recently presented positive effects of class size reduction from a randomized study, but offered three cautions: (a) there is a right way and a wrong way to do it (the wrong way is without professional development and support); (b) reduced class size (which directly effects student/teacher interaction) is a far better statistic than pupil/teacher ratio (which addresses only the number of teachers *in the building*); and (c) more research is needed on the effect of reduced class size. Muller and Schiller (2000), investigating the effects of testing policies on states, ran hierarchical linear modeling on the National Education Longitudinal Study, 88-92 and the National Longitudinal Study of Schools and found inconsistent effects. They interpreted their findings as fully supporting neither side of the testing policy debate, "neither the broad claims of proponents of greater equity nor the grim forecasts of opponents" (p. 211). Even when effective policies are identified, they may be costly to implement. Green (1994) stated that inherent in policy is the difficultly of simultaneously satisfying all needs and maximizing all outcomes. He in fact emphasizes his opinion that greater equity may mean a tradeoff in quality. Evan (1996) noted also that the effect of a change make take so long to appear that time is needed to properly evaluate it. So there appears to be no clear picture of the relationships between demographics or policy on the one hand, and quality of education on the other.

Purpose

Our basic question is to what extent the *Quality Counts 2002* report cards provide valid information to clients and public managers. In this paper, we address two research questions: (a) how comprehensive are the report card grades, i.e., do they appear to cover a range of concerns, or are they highly intercorrelated? This question addresses comprehensiveness. (b) are there environmental factors related to states' grades? This second question implies its converse: is there room for policy to improve grades? This question addresses functionality.

We were interested in looking at variables that define geographic, economic, or demographic climates in the states. Many of these variables cannot be directly controlled by the state or K-12 education. We also wanted to consider what we perceived as important variables that were not necessarily within the control of K-12 education but could be controlled by the state. Through our two research questions, we hope to initiate a discussion about the relationship among *Quality Counts 2002* grades and between specific grades and a limited number of possible predictor variables. Another researcher may have chosen different variables for different reasons, but our purpose was to create a starting point. We intend to continue our analysis and hope that others find ways to refine, critique or expand what we did to stimulate scholarly inquiry and uncover relevant relationships that have implication for policy or at least for discussion of the usefulness of state report cards.

We believe our analysis offers the potential to raise several policy-relevant questions. The impetus for first trying to understand the relationship between report card categories was to see if these relationships were intuitive. If a significant relationship exists, does it make sense? If not, is it possible that the construction of the category needs revision? These are important questions since state-level policymakers are the intended audience of such report cards.

The second line of inquiry focuses on the relationship between select elements of a state's educational context and its report card performance. Strong relationships between a report card grade and demographic factors largely outside state policymaker's control may signify policymakers' inability to affect that grade.

A Review of the Report Card Categories

The brief review of the definition of graded categories contained in the *Quality Counts 2002* report card serves as a summary for the reader. It is not the intent of this article to fully define each subcategory and its mathematical contribution to the category grade. Nor is it our purpose to debate the definition and legitimacy of any category, subcategory or the placement and inclusion of any subcategory until after we completed our analysis. Complete and full descriptions of the K-12 report card categories and subcategories are available in the *Quality Counts 2002* report card (Education Week, 2002).

- Student achievement. The percentage of 4th and 8th graders who score above "proficient" on the year 2000 National Assessment of Educational Progress (NAEP) exam in reading, writing, and math examinations. Several states were missing data for Student Achievement, so we focused on 8th grade scores. We added the three scores of 8th grade performance on reading, writing, and mathematics [g1]for each state. Any state with a missing grade in any of these categories was excluded. Thirty states had complete information on 8th grade performance.
- Standards and accountability. The extent to which a state 1) has developed educational standards, 2) uses various assessment techniques to measure those standards, and 3) holds schools and students accountable for performance.

- Teaching Quality. How well states assess teachers and assure that those teaching are teaching in their field of expertise, and the level of professional support and training for teachers' experiences. Also included as a portion of this grade is the effectiveness of teacher education in the state.
- Resources. This area is broken down into two subcategories: Adequacy and Equity. Each of the two areas had a complete and separate grade, so both were used as K-12 grades. Adequacy measures a state's spending effort on education, including spending per student and percentage of taxable dollars devoted to education. Equity measures a state's effort to equalize per-pupil funding levels across districts.

Method

Our first research question, testing for relationships between report card categories, was fairly straightforward. We conducted a two-tail Pearson correlation analysis among five report card categories from *Quality Counts 2002*. We were primarily interested in looking at the relationships to see if they made intuitive sense. In addition, significant relationships might translate into important policy considerations. For example, a significant negative correlation between two categories might indicate a state's difficulty in obtaining high grades in both of those categories, outlining inherent tradeoffs in state educational emphasis and investment. This would not, however, necessarily call into question the validity of either report card category; it might instead be a sign of the report card's comprehensiveness.

The second research question, which entails using state contextual variables as predictors for the report card categories, was more complex. We were not prepared to state hypotheses. Rather, our intent was to take a step toward developing hypotheses. This meant that our investigation was exploratory and that a regular multiple regression was not appropriate

(McNeil, Newman, and Kelly, 1996). A stepwise regression was the prime methodological candidate for the analysis, and we chose to use backward stepwise regression. Backward stepwise regression takes all independent variables and deletes those that aren't contributing to the goodness of fit. We did not use forward stepwise regression because we did not postulate which variables were most salient (McNeil, Newman, and Kelly).

Since our premise was to test report card categories for the 50 states, the maximum number of observations was fifty. We did not include a regression equation for Standards and Accountability, since we assumed that this is virtually totally a function of state policy making, and unrelated to any contextual varibles. Of the four K-12 report card categories we tested, only achievement did not contain grades for all 50 states, for reasons stated in the previous section. We wanted to outline possible issues that might arise because of our limited observations for Student Achievement. According to the University of Texas Statistics Department website, consensus is lacking regarding the number of observations required per independent variable for multiple regression analysis. The numbers in the literature range from 5 to 50.

Given these considerations, we wanted to at least meet the suggested five-observation minimum, but exceed it where possible. The K-12 report card categories served as the dependent variables, and the various contextual variables that served as our independent predictors all contained 50 observations. For the achievement equation, we did choose 5 predictor variables, placing us at 6 observations per independent variable. The other equations contained fewer predictor variables and contained all 50 observations. Since our intent was purely exploratory, we decided to pursue our analysis even if we were very close to our self-imposed minimum number of observations per independent variable for the achievement equation. As an additional measure, we tracked the statistical significance of the total final

models and each predictive variable contributing to each model. This allowed us to offer cautionary and conservative remarks regarding each model.

Choosing Predictor Variables

In this section, we first offer a general description outlining our selection criteria and rationale for choosing predictor variables. We then list each report card category and outline the associated predictor variables to each category. Our starting point was to review each report card category and discuss whether a measure of income, ethnicity, or both should be included as predictor variables for the regression analysis. It is conceivable that the reader could fashion a counter argument for excluding a particular variable and including another. We encountered some disagreement between ourselves and debated the merits of a list of variables much larger than those we eventually decided upon. The construction of the four equations is given below, and the data sources for each predictor are shown in Appendix B.

Report Card Categories and Predictor Variables

Predictor Variables for Student Achievement: We were interested in whether the often-repeated belief that achievement is just a function of income was true among states with the *Quality Counts 2002* Student Achievement category. We also wanted to test some measure of the financial priority states give to K-12 education (K-12 spending per $1000 of state wealth). We were also interested in some general educational conditions of the state, such as the number of students per teacher, teacher pay, and the percentage of minority students in K-12.

Teacher Quality Predictor Variables: A host of variables could arguably be deemed as influential in teacher quality. Conversations in states around the nation invariably do link teacher quality with some notion of pay (e.g., New Mexico Commission on Higher Education and the New Mexico State Department of Education[gi2], 2000), especially in states were teacher

pay is far below the national average; thus teacher pay was included as a predictor variable. One might argue that large class sizes correlate with teacher quality as well, so the number of students per teacher was the second variable we included in the analysis. Finally, a state's emphasis on K-12 as a priority also seemed like a reasonable variable for inclusion. K-12 spending per $1000 was our final predictor variables for this equation.

Resources Adequacy Predictor Variables: The first issue Resources Adequacy addresses is the level at which each student is funded, and second, how much of a financial priority K-12 is in the state. We thought that income per capita might be a good predictor of how well students and K-12 education is funded. Another initial thought was that states with larger minority populations in their K-12 system and larger class sizes might, for some reason, find it more challenging to adequately fund K-12 education. A large K-12 minority population would indicate a large state minority population. Disparities between minority and majority incomes is well established in the literature (e.g. Phillippe[gi3], 2000, p. 59), so a state with a large minority population might find it more challenging to adequately fund K-12 education.

Resources Equity Predictor Variables: The predictor variables chosen for this equation were the same as for the Adequacy equation, but for different reasons. Attention to K-12 financial equity among districts is largely a function of state philosophy, but we were curious as to whether an indicator of state wealth (income per capita) had an influence on whether equity was actually achieved. In addition, in states like New Mexico, where minority populations are significant, our sense was that there is a stronger inclination toward equitable funding. Thus, we included the measure of the percentage of minority students in the K-12 system. Without any speculations, we also included teacher pay in the analysis, if for no other reason than that we had included it in the Adequacy equation.

Results

The Pearson correlations in Table 1 were obtained from the numerical grades for each report card category for all 50 states. We instructed SPSS to compute the pairwise Pearson correlations. Student achievement was the only category that did not contain 50 observations, as data for 20 states were not available.

\

Put Table 1 about here.

Grade 8 NAEP 2000

Student achievement as measured by Grade 8 NAEP 2000 is the first row in Table 1. One might initially speculate that Teacher Quality should enhance student achievement. Teacher Quality depends on the proportion of those teachers who are certified to teach in their field of expertise and on professional development and the strength of teacher education in the state. If all of these components were strong, it would seem that student achievement would also be strong. We found no correlation between GRADE 8 NAEP 2000 and Teacher Quality, however.

We also found no significant correlation between Grade 8 NAEP 2000 and Standards and Accountability. One interpretation of this is that states may *have* high Standards and Accountability, but it is possible that student achievement is unaffected because some schools *meet* the Standards and Accountability while others simply do not. Without further information, we are left to believe that Standards and Accountability may be in effect in a given state, but the

ability of those Standards and Accountability to leverage improvement or behavioral change at an institutional level is in question.

We did find a significant positive relationship between Grade 8 NAEP 2000 and the Adequacy with which states fund their K-12 systems. Those states that devote greater resources to K-12, spend more per student, and channel a large percentage of tax dollars to K-12 education, tended to do well on Grade 8 NAEP 2000 as well. One indication from this is that money does matter. Perhaps money is not the only answer to improve student achievement, but our analysis suggests that it is part of the answer.

Conversely, Grade 8 NAEP 2000 had a significant negative correlation with Equity. That is, states that do a better job of funding students equally across districts tended to do poorly on Grade 8 NAEP 2000. The relationship between equity and achievement may represent the multiple tradeoffs and policy dilemmas that states will face as they try to find effective K-12 funding solutions. Equal funding to students, regardless of district, is a laudable goal indeed. But are resources spread too thinly when this done? Going further, perhaps equity and student achievement are coincidentally correlated, and any constructed explanation equates to mere speculation at best or misguided assumptions at worst. It appears there are no easy answers.

Standards and Accountability

States with high Standards and Accountability scores tended to do well on Teacher Quality. Neither Standards and Accountability nor Teacher Quality influenced Student Achievement, however. Our supposition for this set of facts is twofold. First, states with high Standards and Accountability have likely set a strong infrastructure in place regarding content specific measurements for student achievement. Teacher Quality, which largely measures subject specific training and certification of teachers, appears to support this infrastructure.

Secondly, and more importantly, the effects of this infrastructure and strong teacher quality may take a few years to manifest in terms of student achievement. This manifestation is not captured in the Student Achievement data reported in the *Quality Counts 2002* report since the student achievement subcategories use 2000 NAEP scores. Standards and Accountability and Teacher Quality capture 2001 activity. In sum, it is probably not fair to conclude that strong Teacher Quality and strong Standards and Accountability do not positively influence performance and student achievement. In line with Evan's (1996) caution, we note that time may be needed to properly evaluate the effect of these factors.

Backward Stepwise Regression Results

Backward stepwise regression considers all predictor variables, for the given dependent variable, in its initial analysis. Each predictor variable is analyzed and an analysis of variance is performed to determine if the complete model (for the given iteration) is significant. The predictor variable with the least significance is removed after each step, and then the analysis is iterated until all variables are stepped out or a significant model is obtained. Any variable whose standard coefficient had a test statistic over .15 was stepped out of our model. This relatively high tolerance was set for analysis purposes, but variables approaching such a limit should be interpreted with caution.

Table 2 summarizes the results for the regression equations tested. Four models were tested, with each K-12 report card category serving as the dependent variable for each model. Included in the table are the multiple correlation coefficients (R) and adjusted coefficients of multiple determination (R squared). We looked at both the correlation between the dependent variable and the entire model (R) and the amount of variance in each report card category explained by the combination of predictor variables in the final model (R squared). The adjusted

coefficients of multiple determination are reported in Table 2 since we reported standardized coefficients for the predictor variables.

Put Table 2 about here.

Student Achievement: The Student Achievement model was significant at the .01 level, despite only having 30 observations. Only 30 states had complete achievement data, by our definition, so all other corresponding cases for predictor variables were excluded. Still, the combination of variables retained in our final model explained 65.7% of the variance. The model, in total, produced a .832 correlation with Student Achievement.

As we suspected, income per capita was retained in the model, and according to the test statistics for the standard coefficient for income per capita, this variable had the strongest effect of all retained variables. K-12 spending per $1,000 of state wealth was also significant. While this measure could be a function of state wealth in general, we interpret K-12 spending per $1,000 of state wealth more as a measure of a state's priority to K-12 education. It is possible for a wealthy state to devote few resources to K-12, or any other state function for that matter. Our interpretation, therefore, is that those states that make K-12 a priority by devoting resources to it, do enhance achievement to some degree.

Interestingly, the percentage of minority students in K-12 negatively contributed to Student Achievement. This is not surprising given the history of correlation between ethnicity and test performance. The implication, therefore, is that states with large minority populations must continue to search for strategies to help such populations achieve, assuming that standardized tests will continue to be the proxy for success.

Teacher Quality: The average number of students per teacher and the priority states give to K-12 were both stepped out of the regression equation for teacher quality. The only variable retained was average pay for teachers, though the percentage of the variance explained was fairly low at 4.7%. Since teacher pay was the only variable retained, the standard coefficient is simply the correlation between teacher pay and Teacher Quality. The model, as a whole, was the weakest of the four that we ran, and the test statistics for average pay indicate that any relationship between teacher pay and Teacher Quality should be interpreted with caution. In one respect, since teacher pay was retained in the analysis, we would cautiously offer that there are many things not included in our model that affect Teacher Quality, but pay, at least to a certain extent, may also bear on that quality.

Resources Adequacy: The measure of priority states give to K-12 was not included as a predictor variable since several of the subcategories that comprise Resources Adequacy directly speak to this issue (e.g. percent of total taxable resources spent on education). For every equation, our intent was to not include any predictor variables that comprised the report card category itself, for it would be a trivial to find a significant relationship between the overall grade and the subcategory that is part of that overall grade.

The model for Resources Adequacy shown in Table 2 explained 22.3% of the variance. The predictor variables that comprised the model yielded a moderate .504 correlation with Resources Adequacy. Only income per capita was stepped out of our model. This was somewhat surprising in that one might guess that wealthy states might more generously fund K-12 education. Perhaps the continued climate of tax reductions and reluctance on the part of policy makers to even speak of raising taxes makes this a somewhat outmoded supposition.

We did find that states with higher teacher pay scored well on Resources Adequacy. It seems logical that states that pay teachers well also fund K-12 education well. Minority enrollment contributed negatively to the model, however. States that have a large percentage of minority K-12 students do not appear to fund the educational systems as well as those states with lower minority student populations. This raises several questions: Is this because states with higher minority enrollments have fewer resources to fund education? Is this because constituents in such states are unable to communicate their needs? Or is this because states simply are not funding school systems that have large numbers of minority students? Obviously, such a finding needs additional investigation to begin uncovering the negative relationship represented by the equation.

Resources Equity: Our Equity model explained 23.3% of the variance and was significant at the .01 level. First, and quite importantly, income per capita produced a negative contribution to the overall model. States with high income per capita tend to be less equitable. That is, there are wider disparities between districts in terms of funding, even though the state tends to be wealthy, relative to other states. This may have to do with anything from state philosophy to how local money is retained and distributed. Even if a state does not allocate local resources, the Equity grade measures how much of an effort that state makes to focus resources on the poorer district. Given this, a conceivable interpretation is that wealthier states do not make as much of an effort to equalize funding across districts.

States with a high percentage of minority students in K-12 and high student to teacher ratios appear to be more successful at equalizing funding across districts. The measure of minority student percentage was retained in the final model. The indication seems to be that states with large K-12 minority populations may work harder at producing equity.

When one looks at the effect of percentage of minority population in K-12 on Equity versus Adequacy, they are the exact opposite. Perhaps this is because working toward equity, or even achieving it, does not imply adequacy. The challenge of adequate funding may well remain even if equity exists.

Discussion

Research question 1(on comprehensiveness)

The answer to our first question is yes, the *Quality Counts 2002* report card grades have comprehensiveness; they are not so highly intercorrelated as to appear all to be measuring essentially the same trait. Of the ten possible correlations, only three were statistically significant, and the highest correlation was a moderate 0.447 between Teacher Quality on the one hand and Standards and Accountability on the other. In fact, as we said above, some of the lack of correlation raises disturbing questions. The near-zero correlation between Standards and Accountability and Teacher Quality on the one hand, and Grade 8 NAEP scores on the other should be monitored to see if, over time, a correlation appears. Many clients and public managers see student achievement data as a kind of pedagogical bottom line, and their interest in Standards and Accountability and Teacher Quality stem purely from their presumed eventual effect on those achievement data.

Furthermore, the possibility that Resources Equity is negatively correlated with other grades raises serious questions. Green's (1994) warnings about policy tradeoffs, cited above, seem relevant here, as our findings suggest that there may be tradeoff between equity and student achievement. It is information like this that must come to the forefront of policy discussions. The issue of equity only initiates the conversation, for the larger solution may encompass

funding amounts or where the funding is focused (e.g. salaries, class size reductions, or teacher preparation and professional development).

Research question 2 (on functionality)

The answer to our second research question is also yes, there are demographic variables associated with the *Quality Counts 2002* grades, but this answer comes with good news and bad news. The good news is that with regard to Teacher Quality, Resources Adequacy, and Resources Equity, demographic factors accounted for only from four to 23 percent of the variance among states. This implies that there is plenty of room for policy to affect the grades. Therefore the *Quality Counts 2002* report cards have functionality. The bad news is with our pedagogical bottom line. Nearly 66% of the between-state variance in Grade 8 NAEP was accounted for by our demographic factors. This leaves comparatively little room for policy effects.

But even here there are opportunities for responses other than fatalism. Along with the claim of Odden, et al. (1995), cited above, that there is little variation in how districts spend money, they recommend that schools begin to spend resources specifically in ways that will produce high achievement. But we also cited sources above that seem to counter the claims of Odden et al. about what expenditures will and will not lead to high achievement. So it is clear that more research and deliberation need to occur in this area. And this is where the *Quality Counts* report cards may not serve us well. Such large data collection projects as *Quality Counts* are seriously constrained by the resources they have to collect data and the availability of those data. The National Center for Public Policy in Higher Education (2000, p. 182) said of its report cards that it restricted itself to "publicly available information that has been collected by government agencies and by nationally recognized private organizations," and it described 15

categories of information that it considered important, but that were unavailable. *Quality Counts* must have faced similar frustrations. It has relied on available data, and sometimes those data are poor proxies for the phenomena of interest. For example, we cited above both Finn's (2002) explanation of why pupil teacher/ratio is a less valuable statistic than class size, and Kennedy's (1999) catalog of the weaknesses of student achievement indicators. Studies aiming to assess validly the effects of policy on education outcomes will need to look for better indicators than the *Quality Counts* grades.

In sum, we see *Quality Counts 2002* as very a useful starter of very important conversations, starter, but we do not believe that it resolves many important questions. McDonnell (1997) called for "rational deliberation" about the meaning of such assessment systems. Clearly this is not a quick fix. She admits that even proposing it leaves her open to the charge of political naïveté. But it is precisely because of the indefiniteness and the lack of guarantees in her prescription that she is in very good company. John Dewey saw the necessary and sufficient criterion for Democracy to be that there are "many interests consciously communicated and shared; and there are varied and free points of contact with other modes of association" (cited in Noddings, 1995, p. 35). We see no solution other than McDonnell's to such ambiguous implications as we have revealed here.

But parties contending over the implications of *Quality Counts 2002* must accept the challenge to do this well. Practitioners must work to overcome their cynicism about policymakers' motives and distance from the problems. Policymakers must eschew the "ugly face" of persuasion, when it becomes "intentionally manipulative, robbing people of their capacity to think independently" (McDonnell, 1997, p. 400). Practitioners must work to understand policymakers' needs and responsibilities to have accountability data, and

policymakers must give up their illusion that accountability data alone will bring about a cheap fix for education so that they can ignore the most costly "capacity building."

Therefore, the development of a state report card is a worthwhile endeavor. And analyses such as ours are necessary to casting a critical eye on report cards and their implications. We see scholars working to refine report cards, and we hope that more states will elect to participate in the data collection. Report cards can promote enlightened discussion. But creation of report cards it is only the beginning of a process. The process must continue with analyses of the report card variables and other associated variables, and with rational open-minded deliberation about the findings of such analyses. The research reported here is one stage in that process.

References

Ackoff, R. (1994). *The dramatic corporation.* New York: Oxford University Press.

Arizona Education Association (2001). Arizona's rank unchanged in "Quality Counts" report, but we've turned the corner. *AEA Advocate.* January/February 2001. p. 13.

Borman, G. D., Stringfield, S.& Rachuba, L. (2000). *Advancing minority high achievement: National trends and promising programs and practices: A report prepared for the National Task Force on Minority High Achievement.* The College Board. Retrieved from the World Wide Web on March 2, 2002 http://www.collegeboard.com/about/association/taskforce/MinorityHighAch.pdf

College Board (1999). *Reaching the top: A Report of the National Task Force on Minority High Achievement. Executive Summary.* Retrieved from the World Wide Web on March 2, 2002 http://www.collegeboard.com/about/association/taskforce/ReachingTheTop.pdf

Cunningham, A. & Wellman J. V. (2001). Beneath the Surface: A Statistical Analysis of the Major Variables Associated with State Grades in *Measuring UP 2000.* Washington, D.C.: The Institute for Higher Education Policy.

Education Week (2002). *The state of the states.* Bethesda, MA: Author

Evan, R. (1996) *The human side of school change.* San Francisco: Jossey-Bass.

Ewell, P. (June 2001). *A review of tests performed on the data in Measuring Up 2000* (National Center Report #01-1) The National Center for Public Policy and Higher Education. San Jose, CA.

Fullan, M. (2001). *The new meaning of educational change* (3rd edition). NY: Teachers College Press.

Finn, J.D. (2002). Small classes in American schools: Research, practice, and politics. *Phi Delta Kappan, 83,* 551-560.

Ginsberg, R. & Berry, B. (1998). The capability for enhancing accountability. In R. J. S. Macpherson (ed.) *The politics of accountability: Educative and international perspectives* (pp. 43-61). Thousand Oaks, CA: Corwin Press.

Green, T. F. (1994). Policy questions: A conceptual study. *Educational Policy Analysis Archives.* Retrieved March 14, 2002 from http://epaa.asu.edu/epaa/v2n7.html.

Gormley, W. T. & Weimer, D. L. (1999). *Organizational report cards.* Cambridge, MA: Harvard University Press.

Kennedy, M. M. (1999). Approximations to indicators of student outcomes. *Educational Evaluation and Policy Analysis, 21*(4), 345-363.

Knapp, M. S. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research (67).* 227-266.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher, 31,* 3-12.

McDonnell, L. M. (1994). Assessment policy as persuasion and regulation. *American Journal of Education, 102,* 394-420.

McNeil, K., Newman, I., & Kelly, F. J. (1996). *Testing research hypotheses with the general linear model.* Carbondale, IL: Southern Illinois University Press.

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing.* New York:Routledge

Muller, C. & Schiller, K.S. (2000). Leveling the playing field? Students' educational attainment and states' performance testing. *Sociology of Education, 73,* 196-218.

National Center for Public Policy and Higher Education. (2000, November). *Measuring up 2000: The state-by-state report card for higher education.* San Jose, California: Author.

New Mexico Commission on Higher Education and New Mexico State Department of Education (2000, September). Teacher supply and demand in New Mexico. K-16 Partnerships to Promote Quality Learning Roundtable.

Noddings, N. (1995). *Philosophy of education.* Boulder, CO: Westview Press.

Oakes, J. with Ormsbeth, T., Bell, R., & Camp, C. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science.* Santa Monica: The RAND Corporation.

Odden, A. Monk, D. Nakib, Yasser, & Picus, L. (1995). The story of the education dollar: No academy awards and no fiscal smoking guns. *Phi Delta Kappan, 77,* 161-168.

Phillippe, K. (Ed.) (2000). *National profile of community colleges: Trends and statistics.* Washington, D.C.: Community College Press.

Richardson, R. C., Reeves-Bracco, K. Callan, P. M. & Finney, J. E. (1999). *Designing state higher education systems for a new century.* American Council on Education: Oryx Press.

Scheurich, J. J., Skrla, L., & Johnson, J. F. (2000). Thinking carefully about equity and accountability. *Phi Delta Kappan, 82*, 293-299.

University of Texas Statistics Department. Statistical Services General FAQ #18: Sample size for multiple regression. Retrieved September 14, 2001 from the World Wide Web: http://www.utexas.edu/cc/faqs/stat/general/gen18.html

Vinovskis, M. A. (1998). *Overseeing the Nation's report card: The creation and evolution of the National Assessment Governing Board (NAGB).* Retrieved from the World Wide Web July 10, 2001. www.nagb.org/pubs/95222.pdf

Wenglinsky, H. (1997). How money matters: The effect of school district spending on academic achievement. *Sociology of Education, 70*, 221-237.

White House (2001). *Remarks by the President on education reform.* Retrieved March 4, 2002 from http://www.whitehouse.gov/news/releases/2001/07/20010705-6.html

Table 1

Pearson correlations (two-tailed tests) among Quality Counts 2002 report card grades

	Grade 8 NAEP, 2000	Standards and Accountability	Teacher Quality	Resources: Adequacy	Resources: Equity
Grade 8 NAEP, 2000		-0.056	0.056	0.407*	-0.446*
Standards and Accountability			0.447**	-0.080	-0.052
Teacher Quality				-0.049	-0.04
Resources: Adequacy					-0.219

** Significant at .001 level

* Significant at .05 level

Table 2

Output from regressions of Quality Counts 2002 grades on indicators of state demographic or economic conditions

	Standardized Regression Coefficients								
	Spending per $1,000 of State Wealth, K-12	Income per Capita (2000)	Average Pay for Teachers (2000)	Percent Minority Students in K-12	Average Number of Students per Teacher	Constant	N	Multiple R	Adjusted Multiple R Squared
Grade 8 NAEP, 2000	0.214	0.667	X	-0.443	X	-15.05	30	0.832	0.657**
Teacher Quality	X		0.257		X	54.82	50	0.257	0.047*
Resources: Adequacy		X	0.367	-0.323		53.529	50	0.504	0.223**
Resources: Equity		-0.396	X	0.363		90.438	50	0.514	0.233**

**Total model significant at .01 level.
* Total model significant at .10 level.
An 'X' indicates variable was in the equation but was stepped out.
A blank cell indicates variable was not included in the equation.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

TM033829

I. DOCUMENT IDENTIFICATION:

Title: The Quality Counts 2002 State Report Cards: Correlations with Measures of State Education Environments

Author(s): Mario C. Martinez and Gary Ivory

Corporate Source: New Mexico State University

Publication Date: April 5, 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Mario C Martinez F. Martinez

Printed Name/Position/Title: Mario C. Martinez, Assist. Prof., Gary Ivory, Assoc. Prof.

Organization/Address: New Mexico State University

Telephone: 505-646-6842 FAX: 505-646-4761

E-Mail Address: mamario@nmsu.edu

Date: 5 April 2002

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to: